

REVISIÓN DE RESULTADOS EXPERIMENTALES EN TÉCNICAS DE PRUEBA Y DE EDUCACIÓN DE CONOCIMIENTOS

Guerini, M.¹, Fernández, E.^{1,2}, Britos, P.^{1,2}, García-Martínez, E.^{1,2}

¹Centro de Ingeniería de Software e Ingeniería del Conocimiento. Escuela de Postgrado. ITBA. Argentina

²Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires. Argentina

{enfernan, pbritos, rgm}@itba.edu.ar

1. Introducción

La cantidad de experimentos realizados en el ámbito de la ingeniería del software ha ido aumentando considerablemente en los últimos años. Estos experimentos cubren un amplio rango de temas, como por ejemplo, técnicas de pruebas en software, técnicas de educación de conocimientos, lenguajes de programación, etc. Sin embargo, si los experimentos no son tratados como investigaciones controladas, formales y rigurosas, existe un riesgo de obtener conclusiones que carezcan de validez.

En este reporte, se analizan 32 experimentos sobre técnicas de prueba y de educación de conocimientos para verificar la correctitud de los mismos.

En la sección 2, se describe el estado de la cuestión, donde se exponen nociones de experimentación y técnicas estadísticas. La sección 3, describe el problema a tratar; y la sección 4, la solución propuesta. Los resultados se exponen en la sección 5. En la sección 6, se presentan las conclusiones, seguidas por las futuras líneas de investigación y la bibliografía en las secciones 7 y 8 respectivamente.

2. Estado de la cuestión.

La experimentación, básicamente, actúa sobre métodos, técnicas y herramientas de la ingeniería del software en ambientes controlados, de forma que se pueda obtener conclusiones acerca de los mismos. El objetivo de la experimentación debería ser el de proveer un marco para calificar, es decir, para poder afirmar, bajo ciertas hipótesis o suposiciones, cuáles son los mejores métodos, técnicas o herramientas en cada caso. También nos va a permitir corroborar empíricamente alguna suposición que haya sido propuesta de forma teórica.

Las técnicas estadísticas cumplen un papel importante en la experimentación ya que permiten interpretar cuantitativamente los datos de los experimentos.

Esta interpretación consta de 3 pasos [33]:

- El primero, es la estadística descriptiva. Mediante este paso se obtienen indicadores de tendencia, como la media, la mediana y la moda. También se obtienen indicadores de dispersión como la varianza, la desviación estándar o el intervalo de variación. Y por último indicadores de dependencia, como por ejemplo, las regresiones lineales o de otro tipo, la covarianza o el coeficiente de correlación.
- El segundo paso es la reducción del conjunto de datos. Los métodos estadísticos correspondientes al tercer paso, todos tienen en común que dependen de la calidad de los datos de entrada. Es decir, que si los datos de entrada son imprecisos, entonces, no importa el método estadístico usado, las conclusiones no van a ser válidas. En este paso, se analizan los datos medidos, para eliminar posibles desviaciones que no son el objeto de estudio. Hay que prestar atención para analizar las causas de eliminar estos datos, de forma de asegurarse que estas desviaciones no formen parte del fenómeno bajo estudio.
- El último de los pasos es el test de hipótesis. En este paso se propone una hipótesis y mediante pruebas estadísticas, se acepta o rechaza la hipótesis planteada. Las pruebas pueden ser paramétricas o no paramétricas. Las paramétricas suponen una cierta distribución

de los datos, mientras que las no paramétricas no tienen esta suposición, y por lo tanto, son más generales. Es decir, siempre que se aplica una técnica paramétrica, se puede aplicar una no paramétrica, pero la inversa no siempre es cierta. Esto es lo que se llama aplicabilidad, es decir, bajo qué condiciones se puede aplicar un tipo de prueba o el otro. El otro punto a tener en cuenta es el Poder. El poder de las pruebas paramétricas es mayor y es por eso que generalmente requieren menor cantidad de datos y por lo tanto experimentos menores.

Los estudios sujetos a análisis en este trabajo, tratan acerca de técnicas de prueba de software y de educación de conocimientos.

Las siguientes técnicas de prueba fueron tratadas en los mismos:

- Técnicas Aleatorias
- Técnicas de Flujo de Control [4]
- Técnicas de Flujo de Datos [3]
- Técnicas de Mutación [34]

Las técnicas de educación estudiadas fueron las siguientes:

- Directas
 - Entrevistas [35]
 - Cuestionarios [35]
 - Educación directa de atributos [25]
- Indirectas
 - Observación de tareas habituales [35]
 - Clasificación de conceptos [10]
 - Análisis de Protocolos [35]
 - Emparrillado [35]
 - Educación de atributos por rangos [25]
 - Descripción ideal de atributos [25]

3. Problema tratado

Este estudio trata de hallar una respuesta a la pregunta de si los experimentos en ingeniería del software son correctamente realizados. Para responder esta pregunta, se analizaron los reportes, verificando los siguientes ítems.

- Correctitud en la aplicación de los procedimientos estadísticos.
- Correctitud en la fundamentación de los resultados.
- Soporte desde el punto de vista estadístico de las conclusiones halladas.

En las siguientes secciones se describe el análisis de estos tres puntos.

4. Solución

La población de casos procede de dos revisiones sistemáticas [36, 37] y consta de un total de 32 estudios, de los cuales 9 provienen del ámbito de las pruebas de software y los 23 restantes del ámbito de la educación de conocimientos.

Estas dos revisiones abarcan experimentos donde el factor humano no tiene una alta incidencia en los resultados (técnicas de prueba), como así también estudios donde la incidencia es alta (educación de conocimientos). [1 a 32]

Sobre cada uno de estos experimentos se analizaron las siguientes cuestiones:

- Presentación de los parámetros de estadística descriptiva
- Correcta reducción del conjunto de datos
- Aplicación de los test de hipótesis (TdH)
- Verificación de los resultados obtenidos

Las tres primeras corresponden a la correctitud de los procesos estadísticos y la cuarta a la fundamentación y soporte estadístico de las conclusiones.

5. Resultados

El análisis de la primer cuestión: presentación de los parámetros de la estadística descriptiva, arrojó los resultados tabulados en la Tabla 1.

Problema en la aplicación del primer paso	Cantidad	%
No publica parámetros estadísticos	7	54%
No publica medida de la dispersión	4	30%
Cálculo erróneo de indicador de dependencia.	1	8%
No publica dispersión ni coeficientes de la regresión efectuada.	1	8%

Tabla 1: Presentación de los parámetros de la estadística descriptiva.

La Tabla 1 muestra que 13 de los 32 estudios presentan falencias con la aplicación del primer paso del procedimiento estadístico.

Al analizar el segundo paso de los procesos estadísticos, se observó que en ninguno de los 32 experimentos se verificó la reducción del conjunto de datos, es decir, que ningún experimento mencionó nada acerca de este paso.

El tercero de los pasos es la aplicación de un test de hipótesis. De los 32 experimentos, en 15 no se aplicó test de hipótesis. En los 17 experimentos restantes se aplicó en test ANOVA, pero en ninguno de estos experimentos se verificaron las hipótesis necesarias para sostener la validez de este test paramétrico.

Finalmente, para estos últimos 15 experimentos, se verificaron las conclusiones halladas. Un 27% de estos estudios no llega a ninguna conclusión, calificando como caso de estudio. Otro 27% no se pudo verificar si las conclusiones eran válidas o no, debido a que las mismas fueron expresadas de forma cualitativa, como, por ejemplo, “la técnica X, es casi tan buena como la técnica Y”. En 3 de los casos se aplicaron pruebas no paramétricas y se verificó que las conclusiones no están soportadas estadísticamente. Dos de los estudios, no fundamentan correctamente las conclusiones, aunque las mismas no pudieron ser verificadas por falta de datos. Finalmente, en otros dos casos, luego de aplicar técnicas no paramétricas, se llegó a la misma conclusión aportada por el estudio. Estos resultados se encuentran resumidos en la Tabla 2.

Resultados	Cantidad	%
No llega a conclusiones.	4	27%
No se pudo verificar	4	27%
No concuerda.	3	20%
No fundamenta las conclusiones.	2	13%
Concuerda.	2	13%

Tabla 2: Verificación de conclusiones

6. Conclusiones

La inspección de la Tabla 3, permite responder las preguntas planteadas:

- Existen reportes de experimentación en software donde los procedimientos estadísticos no son aplicados correctamente.
- Existen reportes donde los resultados no son fundamentados correctamente.
- Existen reportes donde las conclusiones son erróneas.

Paso	Cantidad	Detalle del problema	Cantidad	% en categoría
Estadística descriptiva	41% (13/32)	No publica parámetros estadísticos	22% (7/32)	54% (7/13)
		No publica medida de la dispersión	13% (4/32)	30% (4/13)
		Cálculo erróneo de indicador de dependencia.	3% (1/32)	8% (1/13)
		No publica dispersión ni coeficientes de la regresión efectuada.	3% (1/32)	8% (1/13)
Reducción de conjunto de datos	100% (32/32)	No realiza reducción del conjunto de datos	100% (32/32)	100% (32/32)
Test de hipótesis	47% (15/32)	No aplica test de hipótesis	47% (15/32)	100% (15/15)
Verificación de resultados	47% (15/32)	No llega a conclusiones.	13% (4/32)	27% (4/15)
		No se pudo verificar	13% (4/32)	27% (4/15)
		No concuerda.	9% (3/32)	20% (3/15)
		No fundamenta las conclusiones.	6% (2/32)	13% (2/15)
		Concuerda.	6% (2/32)	13% (2/15)

Tabla 3: Resumen de resultados hallados

4. Formación de Recursos Humanos

En la línea de investigación cuyos resultados parciales se reportan en esta comunicación, se encuentran trabajando: un tesista de maestría de Ingeniería del Software y un tesista de doctorado en Ingeniería Informática.

7. Futuras líneas de investigación

Como futuras líneas de investigación, se propone:

- Ampliar la cantidad de estudios analizados, tomando nuevas revisiones sistemáticas.
- Hacer una evaluación más detallada de los estudios analizando no sólo los parámetros estadísticos, sino también otros aspectos que hacen a la calidad de los mismos.
- Explorar qué medidas deben ser obligatoriamente informadas en un reporte de experimentación en software.

8. Bibliografía

- [1] Mathur A. y Wong W., 1993. *Comparing the fault detection effectiveness of mutation and data flow testing: An empirical study*. Tech. Report SERC-TR-146-P, Software Engineering Research Center.
- [2] Frankl P., Weiss S., Hu C., 1997. *All-uses versus mutation testing: An experimental comparison of effectiveness*. J. Systems and Software, Sept. 1997, 38(3): 235-253.
- [3] E. J. Weyuker. *The Cost of Data Flow Testing: An Empirical Study*, IEEE Transactions on Software Engineering. Vol 16. No 2. February 1990.
- [4] Frankl, P. y Iakounenko, O., 1998. *Further Empirical Studies of Test Effectiveness*. In Proceedings of the ACM SIGSOFT International Symposium on Foundations on Software Engineering, pages 153-162, Lake Buena Vista, Florida, USA.

- [5] Bieman, J. y Schultz, J., 1989. *Estimating the Number of Test Cases Required to Satisfy the All-du-paths Testing Criterion*. ACM. Pages 179-186.
- [6] Offutt, A.J., Rothermel, G. and Zapf, C., 1993. *An Experimental Evaluation of Selective Mutation*. Proceedings of the 15th International Conference on Software Engineering. Pages 100-107. Baltimore, USA. IEEE
- [7] Offutt, A.J. and Lee, S.D., 1994. *An Empirical Evaluation of Weak Mutation*. IEEE Transactions on Software Engineering. Vol. 20(5). Pages 337-344.
- [8] Frankl P., Weiss S., Hu C. *An Experimental Comparison of the Effectiveness of Branch Testing and Data Flow Testing*
- [9] Offutt, A.J. and Lee, S.D. *An Experimental Determination of Sufficient Mutant Operators*
- [10] Eva Hudlicka, et al. *Requirements Elicitation with Indirect Knowledge Elicitation Techniques: Comparison of Three Methods*
- [11] Leonard Adelman. *Measurement Issues in Knowledge Engineering*
- [12] Malcom Eva, et al. *Requirements Acquisition for rapid applications development*
- [13] Victoria Goodrich, Lome Olfman. *An Experimental Evaluation of Task and Methodology Variables For Requirements Definition Phase Success*
- [14] Janete W. Moody, et al. *Enhancing Knowledge Elicitation using the Cognitive Interview*
- [15] J. Michael Moore. *A Comparison of Questionnaire-Based and GUI-Based Requirements Gathering*
- [16] Beth W. Crandall. *A Comparative Study Of Think-Aloud And Critical Decision Knowledge Elicitation Methods*
- [17] Lee A. Freeman. *The effects of concept maps on requirements elicitation and system models during information systems development*
- [18] A. M. Burton, et al. *A Formal Evaluation of Knowledge Elicitation Techniques For Expert Systems*
- [19] R. Schweickert, et al. *Comparing Knowledge Elicitation Techniques: A case Study*
- [20] A. M. Burton, et al. *The Efficacy of Knowledge Elicitation Techniques: a comparison across domains and levels of expertise*
- [21] G. Rugg, et al. *A Comparison of sorting techniques in knowledge acquisition.*
- [22] Glenn Browne, et al. *An Empirical Investigation of User Requirement Elicitation: Comparing the Effectiveness of Prompting Techniques*
- [23] Ritu Agarwal, et al. *Knowledge Aquisition Using Structured Interviewing: An Empirical Investigation.*
- [24] Tino Bech-Larsen, Niels Asger Nielsen, 1999. *A comparison of five elicitation techniques for elicitation of attributes of low involvement products*. Journal of Economic Psychology 20.
- [25] Einar Breivik, Magne Supphellen, 2003. *Elicitation of product attributes in an evaluation context: A comparison of three elicitation techniques*. Journal of Economic Psychology 24.
- [26] Fowles, Salas, et al. *The Utility of Event-Based Knowledge Elicitation*
- [27] Jones, Miles et. al. *The use of a prototype system for evaluating knowledge elicitation techniques*
- [28] Corbridge, C., et al. *Laddering: Technique and tool use in knowledge acquisition*
- [29] Tor J. Larser, et al. *An experimental comparison of abstract and concrete representations in system analysis*
- [30] George M. Marakas, et al. *Semantic Structuring in Analyst Acquisition and Representation of Facts in Requirements Analysis*
- [31] Anna L. Rowe, Nancy J. Cooke ,et al. *Toward an On-Line Knowledge Assessment Methodology: Building on the Relationship Between Knowing and Doing*
- [32] Robert W. Zmud. *The Use of mental Imagery to Facilitate Information Identification in Requirement Analysis.*
- [33] Wohlin C., Runeson P., Höst M., et al., 2000. *Experimentation in Software Engineering. An Introduction*. Kluwer Academic Publishers.
- [34] *The Mothra Software Testing Environment User's Manual*. Technical Report SERC-TR-4-P, Software Engineering Research Center, Purdue University, 1987.
- [35] E. Cordingley, *Knowledge elicitation techniques for knowledge-based systems*, in: D. Diaper (Ed.), *Advances in the Psychology of Human Intelligence*, Ellis Horwood, Chichester, UK, 1989, pp. 89-175.
- [36] Juristo N., Moreno A. y Vegas S. 2003. *Limitations of Empirical Testing Technique Knowledge*. Lecture notes on empirical software engineering archive, pages 1-38. World Scientific Publishing Co., Inc. River Edge, NJ, USA.
- [37] Davis, A.; Dieste o.; Hickey, A.; Juristo, N.; Moreno, A.; 2006; *Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review*; 14th IEEE International Requirements Engineering Conference (RE'06) pp. 179-188